What is claimed is:

1. A method for quantitatively representing objects in a vector space, comprising the steps of:

identifying an object to be processed from a plurality of objects;

extracting a feature corresponding to the object from the plurality of objects;

converting the feature to at least one vector; and

associating the at least one vector with the object.

2. The method of claim 1, wherein the object to be processed comprises a subject document selected from a collection of documents.

3. The method of claim 2, wherein the feature comprises text surrounding the subject document in a host document.

4. The method of claim 2, wherein the feature comprises text represented by the subject document.

5. The method of claim 4, wherein the converting step comprises the steps of:

identifying each unique word within the text represented by all documents in the collection of documents;

counting the occurrences of each unique word in the subject document; and

creating a vector having a number of dimensions equal to the number of unique words in the collection of documents, and further having as each element a numeric value representative of the number of occurrences in the subject document of the corresponding word.

6. The method of claim 5, wherein the value representative of the number of occurrences in the subject document of the corresponding word is calculated as the token

frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word.

7. The method of claim 2, wherein the feature comprises the subject document URL representing the subject document in the collection of documents.

8. The method of claim 7, wherein the converting step comprises the steps of:

identifying each unique word within the URLs representing all documents in the collection of documents; and

counting the occurrences of each unique word in the subject document URL;

creating a vector having a number of dimensions equal to the number of unique words in the URLs representing all documents in the collection of documents, and further having as each element a numeric value representative of the number of occurrences in the subject document URL of the corresponding word.

9. The method of claim 8, wherein the value representative of the number of occurrences in the subject document URL of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word.

10. The method of claim 2, wherein the feature comprises inlinks in the collection of documents linking to the subject document.

11. The method of claim 10, wherein the converting step comprises the steps of:

identifying each document having links within the collection of documents;

determining how many times each document having links points to the subject document; and

creating a vector having a number of dimensions equal to the number of documents having links in the collection of documents, and further having as each

element a numeric value representative of the number of links in each corresponding document linking to the subject document.

12. The method of claim 11, wherein the numeric value representative of the number of links in each corresponding document linking to the subject document is calculated as the token frequency weight of the corresponding link multiplied by the inverse context frequency weight of the corresponding link.

13. The method of claim 10, wherein the converting step comprises the steps of:

identifying each document having hyperlinks within the collection of documents, and further identifying each unique word associated with URLs defining hyperlinks in each document;

counting the occurrences of each unique word in the URLs defining hyperlinks pointing to the subject document; and

creating a vector having a number of dimensions equal to the number of unique words associated with URLs defining hyperlinks within the collection of documents, and further having as each element a numeric value representative of the number of occurrences in the URLs defining hyperlinks pointing to the subject document of the corresponding word.

14. The method of claim 13, wherein the numeric value representative of the number of occurrences in the URLs defining hyperlinks pointing to the subject document of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word.

15. The method of claim 2, wherein the feature comprises outlinks in the subject document linking to other documents.

16. The method of claim 15, wherein the converting step comprises the steps of:

identifying each other document linked to by all documents within the collection of documents; and

creating a vector having a number of dimensions equal to the number of other documents linked to by documents in the collection of documents, and further having as each element a numeric value representative of the number of links in the subject document linking to each corresponding other document.

17. The method of claim 16, wherein the numeric value representative of the number of links in the subject document linking to each corresponding other document is calculated as the token frequency weight of the corresponding link multiplied by the inverse context frequency weight of the corresponding link.

18. The method of claim 15, wherein the converting step comprises the steps of:

identifying each unique word associated with URLs defining hyperlinks in each document in the collection of documents;

counting the occurrences of each unique word in the URLs defining hyperlinks in the subject document; and

creating a vector having a number of dimensions equal to the number of unique words associated with the URLs defining hyperlinks in each document, and further having as each element a numeric value representative of the number of occurrences in the URLs defining hyperlinks in the subject document of the corresponding word.

19. The method of claim 18, wherein the numeric value representative of the number of occurrences in the URLs defining hyperlinks in the subject document of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word.

20. The method of claim 2, wherein the feature comprises the genre of the text represented by the subject document.

21. The method of claim 20, wherein the converting step comprises the steps of:

for each possible text genre, processing the subject document to calculate the probability that the subject document is of the corresponding genre; and

creating a vector having a number of dimensions equal to the number of possible text genres, and further having as each element a numeric value representative of the probability that the subject document is of the corresponding genre.

22. The method of claim 2, wherein the feature comprises the color histogram for an image represented by the subject document.

23. The method of claim 22, wherein the converting step comprises the steps of:

quantizing the image represented by the subject document into a multi-dimensional color model;

creating a color histogram having a plurality of bins for each dimension in the color model, each bin corresponding to a unique combination of binary bits representing information from the associated dimension of the color model;

counting each of a plurality of pixels from the image in a corresponding bin associated with each dimension of the color model; and

creating a vector having a number of dimensions equal to the total number of bins in the color histogram, and further having as each element a numeric value representative of the number of pixels in the image corresponding to the corresponding histogram bin.

24. The method of claim 23, wherein the plurality of pixels from the image in the counting step comprises all of the pixels in the image.

25. The method of claim 24, wherein the plurality of pixels from the image in the counting step comprises an approximately uniformly spaced set of subsampled pixels from the image.

26. The method of claim 23, wherein:

the color model comprises a three-dimensional hue, saturation, and value color model;

each dimension of the color model is represented by two bits of information; and

the color histogram has four bins for each dimension in the color model, for a total of twelve bins.

27. The method of claim 23, wherein the image represented by the subject document comprises a region of a bitmap.

28. The method of claim 2, wherein the feature comprises the color complexity of an image represented by the subject document.

29. The method of claim 28, wherein the converting step comprises the steps of:

quantizing the image represented by the subject document into a multi-dimensional color model;

determining the maximum number of pixels in any row in any image represented by a document in the collection of documents;

determining the maximum number of pixels in any column in any image represented by a document in the collection of documents;

creating a horizontal complexity histogram and a vertical complexity histogram, each having a number of bins equal to the maximum number of pixels in any row and in any column, respectively;

identifying horizontal runs of pixels of all possible lengths in the quantized image, and for each possible length, counting the number of pixels in a plurality of rows of the

quantized image belonging to the horizontal runs in a corresponding bin of the horizontal complexity histogram;

identifying vertical runs of pixels of all possible lengths in the quantized image, and for each possible length, counting the number of pixels in a plurality of columns of the quantized image belonging to the vertical runs in a corresponding bin of the horizontal complexity histogram;

creating a horizontal complexity vector having a number of dimensions equal to the maximum number of pixels in any row, and further having as each element a numeric value representing the number of pixels in the image in the corresponding horizontal histogram bin; and

creating a vertical complexity vector having a number of dimensions equal to the maximum number of pixels in any column, and further having as each element a numeric value representing the number of pixels in the image in the corresponding vertical histogram bin.

30. The method of claim 29, wherein the plurality of rows comprises all rows of the quantized image, and wherein the plurality of columns comprises all columns of the quantized image.

31. The method of claim 29, wherein the plurality of rows comprises an approximately uniformly spaced set of subsampled rows from the image, and wherein the plurality of columns comprises an approximately uniformly spaced set of subsampled columns from the image.

32. The method of claim 29, wherein:

the color model comprises a three-dimensional hue, saturation, and value color model; and

each dimension of the color model is represented by two bits of information.

33. The method of claim 29, further comprising the step of concatenating the horizontal complexity vector and the vertical complexity vector to form a complexity

vector having a number of dimensions equal to the maximum number of pixels in any row plus the maximum number of pixels in any column.

34. The method of claim 28, wherein the converting step comprises the steps of:

quantizing the image represented by the subject document into a multi-dimensional color model;

determining the maximum number of pixels in any row in any image represented by a document in the collection of documents;

determining the maximum number of pixels in any column in any image represented by a document in the collection of documents;

creating a horizontal complexity histogram and a vertical complexity histogram, each having a selected number of bins corresponding to a plurality of quantized ranges of run lengths;

identifying horizontal runs of pixels of all possible lengths in the quantized image, and for each possible length, counting the number of pixels in a plurality of rows of the quantized image belonging to the horizontal runs in a corresponding bin of the horizontal complexity histogram;

identifying vertical runs of pixels of all possible lengths in the quantized image, and for each possible length, counting the number of pixels in a plurality of columns of the quantized image belonging to the vertical runs in a corresponding bin of the horizontal complexity histogram;

creating a horizontal complexity vector having a number of dimensions equal to the selected number of bins in the horizontal complexity histogram, and further having as each element a numeric value representing the number of pixels in the image in the corresponding horizontal histogram bin; and

creating a vertical complexity vector having a number of dimensions equal to the number of bins in the vertical complexity histogram, and further having as each element a numeric value representing the number of pixels in the image in the corresponding vertical histogram bin.

35. The method of claim 34, wherein:

a bin $b_x$ in the horizontal complexity histogram corresponding to a horizontal run of length $r_x$ is identified by a relationship $b_x = \text{floor}(r_x(N\text{-}1) / (n_x/4)) + 1$, where $N$ is the selected number of bins in the horizontal complexity histogram and $n_x$ is a maximum number of pixels in any row of an image in the collection; and

a bin $b_y$ in the vertical complexity histogram corresponding to a vertical run of length $r_y$ is identified by a relationship $b_y = \text{floor}(r_y(N\text{-}1) / (n_y/4)) + 1$, where $N$ is the selected number of bins in the horizontal complexity histogram and $n_y$ is a maximum number of pixels in any row of an image in the collection.

36. The method of claim 34, wherein the plurality of rows comprises an approximately uniformly spaced set of subsampled rows from the image, and wherein the plurality of columns comprises an approximately uniformly spaced set of subsampled columns from the image.

37. The method of claim 34, wherein:

the color model comprises a three-dimensional hue, saturation, and value color model; and

each dimension of the color model is represented by two bits of information.

38. The method of claim 34, further comprising the step of concatenating the horizontal complexity vector and the vertical complexity vector to form a complexity vector having a number of dimensions equal to the selected number of bins in the horizontal complexity histogram plus the selected number of bins in the vertical complexity histogram.

39. The method of claim 1, wherein the object to be processed comprises a subject user selected from a user population.

40. The method of claim 39, wherein the feature comprises the documents in a collection of documents accessed by the subject user.

41. The method of claim 40, wherein the converting step comprises the steps of:

identifying each unique document in the collection of documents;

calculating the number of times the subject user accessed each document in the collection of documents; and

creating a vector having a number of dimensions equal to the number of documents in the collection of documents, and further having as each element a numeric value representative of the number of times the subject user has accessed the corresponding document.

42. The method of claim 41, wherein the value representative of the number of times the subject user has accessed the corresponding document is calculated as the token frequency weight of the corresponding document multiplied by the inverse context frequency weight of the corresponding document.